DOCUMENT RESUME

ED 450 158 TM 032 348

AUTHOR	McMillan, James H.; Lawson, Sonya R.
TITLE	Secondary Science Teachers' Classroom Assessment and Grading Practices.
SPONS AGENCY	Metropolitan Educational Research Consortium, Richmond, VA.
PUB DATE	2001-01-09
NOTE	35p.
PUB TYPE	Reports - Research (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Academic Achievement; Educational Assessment; Educational Practices; *Grading; *Science Teachers; Secondary Education; *Secondary School Teachers; *Student Evaluation; *Teacher Attitudes

ABSTRACT

This study investigated the assessment and grading practices of 213 secondary science teachers representing urban, suburban, and rural schools. Teachers indicated the extent to which they used various factors in grading students, the types of assessments used, and the cognitive level of these assessments. The results indicate a wide variation in practices. Teachers appear to conceptualize six major factors in grading students, placing greatest weight on academic performance and academic-enabling behaviors, such as effort and improvement, and much less emphasis on external benchmarks, extra credit, homework, and participation. Factor analysis for types of assessments used resulted in four components: constructed response assessments, assessment development, objective assessments, and major examinations. In terms of cognitive level of assessments, teachers differentiated between recall and higher-order thinking skills. However, there were few relationships among these components and grade level. With respect to ability level of the class, teachers of higher ability students tended to use types of assessments, cognitive levels of assessments, and grading criteria that mirrored those encouraged by recent literature, such as the use of performance assessments. Teachers of low ability students, in contrast, emphasized recall knowledge and graded homework, and focused less on academic achievement and higher order thinking. The results are discussed in light of other research indicating that teachers use a "hodgepodge" of factors when assessing and grading students. (Contains 6 tables and 18 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 450 158

J. H. McMillan

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Secondary Science Teachers' Classroom Assessment and Grading Practices

James H. McMillan
Sonya R. Lawson
Virginia Commonwealth University

January 9, 2001

TM032348

Funds to support this research were provided by the Metropolitan Educational Research Consortium, Virginia Commonwealth University. The findings and conclusions are those of the authors and not members of the Consortium.

Abstract

Secondary Science Teachers' Classroom Assessment and Grading Practices

This study investigated the assessment and grading practices of 213 secondary science teachers representing urban, suburban, and rural schools. Teachers indicated the extent to which they used various factors in grading students, the types of assessments used, and the cognitive level of these assessments. The results indicate a wide variation in practices. Teachers appear to conceptualize six major factors in grading students, placing greatest weight on academic performance and academic-enabling behaviors, such as effort and improvement, and much less emphasis on external benchmarks, extra credit, homework, and participation. Factor analysis for types of assessments used resulted in four components: constructed-response assessments, assessment developer, objective assessments, and major examinations. In terms of cognitive level of assessments, teachers differentiated between recall and higher-order thinking skills. However, there were few relationships among these components and grade level. With respect to ability level of the class, teachers of higher ability students tend to use types of assessments, cognitive levels of assessments, and grading criteria that mirror that encouraged by recent literature such as use of performance assessments. Teachers of low ability students, in contrast, emphasize recall knowledge and graded homework and focus less on academic achievement and higher order thinking. The results are discussed in light of other research indicating that teachers use a "hodgepodge" of factors when assessing and grading students.

A significant amount of recent literature has focused on classroom assessment and grading as essential aspects of effective teaching. There is an increased scrutiny of assessment as indicated by the popularity of performance assessment and portfolios, newly established national assessment competencies for teachers (American Federation of Teachers, National Council on Measurement in Education, and National Education Association, 1990), and the interplay between learning, motivation, and assessment (Brookhart, 1993, 1994; Tittle, 1994). Previous research documents that teachers tend to award a "hodgepodge grade of attitude, effort, and achievement" (Brookhart, 1991, p. 36). It is also clear that teachers use a variety of assessment techniques, even if established measurement principles are often violated (Frary, Cross, & Weber, 1993; Plake & Impara, 1993; and Stiggins & Conklin, 1992).

Given the variety of assessment and grading practices in the field, the increasing importance of assessment, the critical role each classroom teacher plays in determining assessments and grades, and the trend toward greater accountability of teachers with state assessment approaches that are inconsistent with much of the current literature, there is a need to fully understand current assessment and "hodgepodge" grading practices. The research literature on classroom assessment practices of secondary science teachers shows some trends, but there are limitations in the nature of the research that restrict a more complete understanding of these practices, such as the use of small convenient samples, instrumentation, and the lack of consideration of ability levels of classes. The purpose of this study was to describe the classroom assessment and grading practices of secondary

science teachers, and determine if meaningful relationships exist between these practices and grade level and ability levels of different classes.

Assessment Practices

Several researchers have examined secondary teachers' classroom assessment practices. Stiggins and Bridgford (1985) asked 228 teachers to describe their classroom assessment practices in terms of use, preferences, attitudes, and role of performance assessment. Across grade levels, teacher-made objective tests and structured performance tests gradually increase in importance whereas reliance on published and spontaneous performance tests declines. Science teachers appear to pace more emphasis on their own objective tests. Sixty-eight percent of the teachers reported using structured performance assessments in their classrooms.

In a review of research studies concerning teachers' assessment practices, Marso and Pigge (1993) also found that teachers use primarily self-constructed assessments. Science teachers relied on traditional paper and pencil tests more so than English, history, and social studies teachers. They concluded that a variety of assessment formats were used by these teachers. Consistent with this study, Gullickson (1985) surveyed 50 science seventh and tenth grade teachers and found that teachers relied most on teacher-made assessments. Seventh grade teachers used papers, essays, and discussion more often than tenth grade teachers, and science teachers used papers, essays, and discussion less often than teachers of other subjects. Science teachers also used more objective assessments. In addition, Fray et al. (1993) surveyed 536 secondary teachers and found that objective assessments were used most, followed by projects, term papers, and essays.

Lawrenz and Orton (1989) asked 285 seventh and eighth grade science and mathematics teachers to describe their emphasis on objectives, assessment categories, and assessment items. Science teachers had more variety in their assessment categories and gave more emphasis to class discussion, attendance, behavior, and projects than mathematics teachers. Science teachers were more likely to emphasize true-false, multiple choice, and essay type items, and placed more emphasis on items that required the definition of concepts, that required students to explain their reasoning, and that had more than one answer. Science teachers reported a strong belief in using hands-on experiences.

Bol and Strage (1996) interviewed ten high school biology teachers and reviewed their course documents. While teachers wanted their students to develop higher-order thinking skills, their assessment practices did not support these goals. Specifically, 50% of the items required only basic knowledge, while almost none required application. Interviews with these teachers revealed that they were not aware of this contradiction. In an ethnographic investigation of 15 high school science teachers, Gallagher and Tobin (1987) found that teachers equate task completion with student learning and emphasis is placed more on rote memorization of factual information than on comprehension, applications and processes of science. Also, they found that teachers offered "watered-down versions' of regular class material to unmotivated and poor achieving students.

Finally, Stiggins and Conklin (1992) asked 24 teachers to keep a journal on their classroom assessment practices. Teachers were found most interested in assessing student mastery or achievement, and performance assessment was used frequently. The

nature of the assessments used in each class was coupled closely with the roles each teacher set for their students.

Grading Practices

A number of studies have investigated teachers' grading practices. From a survey of seventh and tenth grade teachers, Gullickson (1985) found that science teachers relied heavily on teacher-made objective tests, but also used citizenship and participation in class to determine course grades. A study by Stiggins, Frisbie, and Griswold (1989) provided an analysis of grading practices as related to recommendations of measurement specialists and newly established Standards for Teacher Competence in Educational Assessment of Students (American Federation of Teachers, National Council on Measurement in Education, National Education Association, 1990). In this study, the authors interviewed and/or observed 15 teachers on 19 recommendations from the measurement literature. They found that teachers use a wide variety of approaches to grading, and that they wanted their grades to reflect fairly both student effort and achievement. They also wanted the grades to motivate students. Contrary to recommended practice, it was found that teachers valued student motivation and effort, and they set different levels of expectation based on student ability.

Brookhart (1993) investigated the meaning teachers give to grades and the extent to which value judgments are used in assigning grades. Eighty-four teachers responded to a questionnaire with multiple choice and open-ended questions. The results indicated that low ability students who tried hard would be given a passing grade even if the numerical grade were failure, while working below ability level did not affect the numerical grade. An average or above average student would get the grade earned, whereas a below

average student gets a break if there is sufficient effort to justify it. Teachers were divided about how to factor in missing work. About half indicated that a zero should be given, even if that meant a failure for the semester. The remaining teachers would lower the grade but not to a failure. The teachers' written comments showed that they strive to be "fair" to students. Teachers also seemed to indicate that a grade was a form of payment to students for work completed. More comments indicated that grades were something students earned as compared to grades indicating academic achievement, as compensation for work completed. This suggests that teachers, either formally or informally, include conceptions of student effort in assigning grades. Because teachers are concerned with student motivation, self-esteem, and the social consequences of giving grades, using student achievement as the sole criteria for determining grades is rare. This is consistent with earlier work by Brookhart (1991), in which she pointed out that grading often consists of a "hodgepodge" of attitude, effort, and achievement. A limitation of this study is the small sample of teachers and the use of only three nonachievement factors in scenarios that subjects responded to (effort/ability, missing work, and improvement). In addition, the subjects in the study were taking a university measurement course, which could result in socially desirable responses or answers that reflect the perspectives of the instructor.

Feldman and Alibrandi (1998) also report findings concerning the "hodgepodge" nature of assigning grades. Ninety-one high school science teachers responded to a survey about types of assessments used, weight given each assessment, and the mechanism used to determine student's grades. Interviews were also conducted. Half of the teachers (50%) reported they based student's grades on achievement, 28% used

comparable students, 16% used individual student ability and 2% used students' growth during the course. Project work, major examinations, performance assessments, portfolios, journals or oral examinations were rarely used. Fray et al. (1993) obtained similar results from 536 secondary teachers from all academic subjects. More than two thirds of the teachers agreed that ability, effort, and improvement should be included in determining grades. Cizek, Fitzgerald, and Rachor (1996) reported similar findings regarding the "hodgepodge" nature of grading. Almost all teachers used formal achievement measures in grading, attendance, ability, participation, demonstration of effort, conduct, and at least half of the teachers used other "achievement-related" factors.

Several limitations of current research exist. One is that the studies do not differentiate grading practices by ability level of the classes. Further research needs to be done to evaluate how ability level influences the type of assessments teachers use and the cognitive level of those assessments. Also, several studies measure teacher beliefs instead of their actual practices (e.g., Brookhart, 1991; Frary et al., 1993; Feldman and Alibrandi, 1998; Stiggins and Bridgeford, 1985; and Lawrenz and Orton, 1989). Another limitation is that the factors used to determine grades have been considered separately. Only one study, Fray et al. (1993), grouped the factors into meaningful categories to analyze their joint effect.

The present study used a relatively large sample of secondary science teachers (grades 6-12) to describe assessment and grading practices in a way that builds upon and extends previous studies, with methods to address weaknesses in prior studies. The critical role of effort and other non-achievement factors in grading is examined, as is the way these different factors cluster together in describing teachers' practices. It was

designed to document differences in actual assessment and grading practices conducted for a specific class taught by each teacher. Four specific research questions were addressed:

1) What is the current state of assessment practice and grading by secondary teachers?
2) What are major types of assessment, grading factors, and cognitive level of assessments that are used by secondary teachers?
3) How do types of assessment, factors used in grading, and cognitive level of assessments cluster into meaningful components?
4) What are the relationships between grade level, ability level of the class, and assessment and grading practices?

Methodology

Sample

The population included 261 grade 6-12 regular classroom science teachers from 69 schools in seven urban/metropolitan Virginia school districts. Completed surveys were returned by 213 teachers from 58 schools (96 middle and 117 high school). The response rate by school was 84%, and, by teachers, it was 89%.

Instrument

A questionnaire, consisting of closed-form items, was used to document the extent to which teachers emphasized different assessment and grading practices. A six point scale, ranging from *not at all* to *completely*, was constructed to allow teachers to indicate usage without the constraints of an ipsative scale that is commonly used in this area (e.g., percentage each factor contributes to grades). Also, the questions were worded to obtain

information about actual teacher practices in relation to a specific class of students, rather than about global teacher beliefs. This was done to provide a more focused point of reference for the teachers that would allow comparisons of different kinds of classes. Teachers were asked to indicate, for the most typical class they taught, the subject matter of the class, the grade level of the class, and the ability level of the class (honors, AP, standard, remedial). The stem for the items was the following:

To what extent were final first semester grades of students in your single class described above based on:

The initial set of items was drawn from previous questionnaires that had been reported in the literature, as well as research on teachers' assessment and grading practices (Frary et al., 1993; Stiggins & Conklin, 1992; Brookhart, 1994). The items included factors that teachers consider in giving grades, such as student effort, improvement, academic performance, types of assessments used, and the cognitive level of the assessments (e.g., knowledge, application, reasoning. Content-related evidence for validity for the initial draft of 47 items was strengthened by asking 15 teachers to review the items for clarity and completeness of covering most if not all assessment and grading practices used. Appropriate revisions were made to the items, and a second pilot test with a school division outside of the sample was used to gather additional feedback on clarity, relationships among items, item response distributions, and reliability. Twenty three teachers participated in the second pilot test. Item statistics were used to reduce the number of items to 27. Items that showed a high correlation or minimum variation were eliminated, as well as items that were weak in reliability. Reliability was assessed by asking the teachers in the second pilot test to retake the questionnaire following a four

week interval. The stability estimate was done by examining the percentage of matches for the items. Items that showed an exact match of less than 60% were deleted or combined with other items. The revised questionnaire included 34 items in the three categories (19 items assessing different factors used to determine grades, 11 items assessing different types of assessments used, and 4 items assessing the cognitive level of the assessments). The average exact match for the items was 46% of the teachers; 89% of the matches were within one point on the six point scale.

Procedure

The surveys were completed in early February, soon after the end of the first semester. School division central administrators communicated to teachers that the questionnaire was to be completed, and were responsible for distribution and collection. The questionnaire took about 15 minutes to complete. Teachers were assured that their responses would be confidential. There was no information was on the form that could be used to identify the teachers.

Data Analysis

The data analyses were primarily descriptive, using frequencies, percentages, means, medians, standard deviations, and graphic presentations to summarize overall findings and trends. An exploratory factor analysis was used to reduce the number of variables investigated within each of the three categories of items. Relationships between assessment and grading practices used by the teachers and cognitive levels of assessments, and grade level and ability level of the classes, were examined through analysis of variance procedures.

Findings

The descriptive results are presented first, followed by the results of data reduction procedures, and relationships between assessment and grading practices and cognitive levels of the assessments, and grade level, and ability level of the class. Table 1 shows the number of classes broken out by grade level and ability level.

Descriptive Results

The means and standard deviations for factors used to determine grades, types of assessments, and the cognitive level of the assessments are reported in Table 2. Table 3 shows the raw score frequency distributions of a few questions to illustrate the spread of scores across the different points in the scale.

For this group of science teachers as a whole, there were some factors that contribute very little, if anything, to grades (means below 2): disruptive student performance, grade distributions of other teachers, performance compared to students from previous years, school division policy about the percentage of students who may obtain different grades, and extra credit for nonacademic performance. Also, a few factors clearly contribute most, ranging from "quite a bit" to "extensively" (means above 4): academic performance as opposed to other factors, performance compared to a set scale of percentage correct, and specific learning objectives mastered.

Five factors were used to at least "some" extent to determine grades (means at or above 3): student effort, ability levels of students, quality of graded homework, degree to which student pays attention and participates in class, and inclusion of zeros for incomplete assignments. There was a fairly large standard deviation reported for these items, showing considerable variation in the extent to which the factors were used for

grading. For example, the mean for student effort was 3.25, with a standard deviation of 1.09. By examining the frequency distribution for this question in Table 3, approximately 40% of the teachers responded "quite a bit," "extensively," or "completely". About 20% of the teachers indicated "not at all" or "very little" to using student effort. This represents a considerable difference among these teachers in the extent to which effort is included in grading. The same kind of variation occurs with other items that tend to average in the middle of the scale.

Concerning types of assessments used there is great reliance on assessments designed primarily by the science teachers themselves, with relatively little reliance on those provided by publishers (see Table 2). Objective assessments are used more frequently than essay type questions, though not by a large margin (means of 4.03 and 3.22, respectively). There is considerable use of performance assessments and individual student projects. Oral presentations and authentic assessments are used least. The standard deviations with respect to types of assessments (about 1 point on the scale) point to considerable variation.

Regarding the cognitive levels of the assessments, student understanding was rated highest, with a strong emphasis on both reasoning and application. Recall knowledge was used least. It is interesting to note that a high percentage of the teachers indicated that they use assessments measuring recall knowledge quite a bit (39%), extensively (7%), or completely (2%). While the percentages for measuring student understanding were higher (47%, 33%, 3%, respectively), it appears that for many of the teachers there was nearly as much emphasis at the recall level as at understanding.

Data Reduction

Three factor analyses, using varimax rotation, were used to reduce the items to fewer, more meaningful, components. One was for factors used in grading (19 items), one for types of assessments (11 items), and one for cognitive levels of assessments (4 items). The results of these analyses are summarized in Table 4.

The factor analysis for items used in grading resulted in six components (grading 1-6) with eginvalues greater than 1. The first component was comprised of four items that emphasized student effort, ability, and improvement. These items could be considered enablers to academic performance, important indicators to teachers to judge the degree to which the students had tried to learn, and by implication, actually learned. The second component loaded on four items that included external benchmarks (comparisons with other students, and grade distributions of other teachers). A third component loaded highly on the use of extra credit and borderline cases. A fourth component loaded on three items focusing on academic achievement of the student (performance and learning objectives mastered). The fifth and sixth components consisted of items describing student attention and participation in class and quality of completed homework, respectively.

The factor analysis for types of assessments used resulted in four components (types 1-4). The first component was comprised of six items that described some kind of constructed-response assessments, such as essay-type questions, performance-based, and projects. The second component included two items that focused on how assessments are constructed (by publisher or teacher-made). The third component loaded highly on one items concerning objective assessments and performance quizzes. The fourth

component loaded on one item regarding major exams. Finally, with respect to cognitive level of assessments, measuring higher-order thinking (understanding, reasoning, and application) formed the first component (level 1) and recall knowledge formed the second component (level 2).

Relationship Results

Twelve one-way ANOVAs were performed, with Sheffe follow up tests, to examine the relationship between the twelve component scores and grade level. Table 5 shows that statistically significant differences were found with only three components. None of the components representing factors used in grading showed significant relationships. With respect to type of assessments used, only one component, major exams, related to grade level. A clear trend was found in the use of major exams, in that high school teachers used major exams significantly more than middle school teachers. Neither component that identified the cognitive level of assessment showed statistically significant differences between grade levels.

The relationship between the twelve component scores and ability level of the class was studied using univariate ANOVAs. A significant difference was found with five components (Table 6). Trends were found across ability levels with all components. Academic achievement was emphasized most in advanced/AP classes, less in standard classes and least in basic/remedial classes. For component six, the same pattern was found, in that the quality of graded homework was used more often in advanced/AP and standard classes than in the basic/remedial classes. In terms of types of assessments used, major exams were emphasized more in the advanced/AP courses than either standard or basic/remedial courses. Regarding the cognitive level of assessments used, it

was found that teachers in advanced/AP courses stressed higher-order thinking more than teachers in standard and basic classes. In addition, teachers in basic/remedial classes emphasized recall knowledge more than either standard or advanced/AP classes.

Discussion

The results of the analyses were consistent with the findings from earlier research by Brookhart (1994), Feldman and Alibrandi (1998), Lawrenz and Orton (1989), Fray et al. (1993), and Cizek et al. (1996) and show that most secondary science teachers use a variety of factors in grading students. There appears to be six conceptually meaningful variables that secondary science teachers use when grading students: effort ability and improvement, external benchmarks, extra credit and borderline cases, academic achievement, participation, and graded homework. Given the relatively low emphasis on comparisons with other students, extra credit, and the infrequent occurrence of borderline cases, these results suggest that teachers conceptualize two major ingredients: academic achievement and effort, ability, and improvement. Of these two, clearly academic achievement is most important, as also reported by Stiggins and Conklin (1992) and Feldman and Alibrandi (1998), but the results of the present study show that academic-enablers, such as effort, participation, and improvement, are also very important for many teachers. Frary et al. found that teachers in their study believed that extraneous factors such as effort and ability should influence grades. Only one nonachievement trait (effort) was reported as being used in a small way for assigning grades in the study conducted by Feldman et al. (1998). Lastly, in a study of seventh and eighth grade science teachers, Lawrenz and Orton (1989) found an emphasis on behavior in the assessment practices and use of a variety of tools to determine student grades.

The use of external benchmarks, such as performance compared to other students, was used very little by most teachers. Extra credit for nonacademic performance is not used very often, but teachers do tend to use extra credit for academic performance, a separate factor closely linked to nontest indicators for borderline cases. Also, it was found that nongraded homework is a separate component. This suggests the use of graded homework, may be a practice distinct from the use of nongraded homework. Lastly, participation is used quite often as a factor in determining student grades and appears as an individual component. This is surprising in that it would appear to be an academic enabler, but yet is not included as a factor in that component.

Disruptive student behavior, grade distributions of other teachers, and norm-referenced interpretations contribute little to grading. However, some kind of norm-referencing is used by many science teachers, as shown by the factors included in the external benchmark component. This is surprising in that all the districts involved in the study have criterion-referenced grading scales. This suggests that teachers need to use some sort of comparative data. A large number of teachers include zeros for incomplete assignments as a factor in grading. Due to the variety of methods of including zeros in grade calculations, this suggests a need to explore in depth how calculating zeros is accomplished.

This study reveals much variation in the types of factors secondary science teachers' use in determining grades, with relatively little difference between teachers at different grade levels. This suggests that teachers differ on how they weigh these factors. This is comparable to findings by Cizek et al. (1996) that grading practices are quite different among teachers and suggests that the meaning of grades conveyed to students

and parents vary greatly. Future research needs to be conducted to examine why there is so much variation among teachers' grading practices. Those teachers that emphasize effort may be sending a message to the students and parents that they demonstrate an adequate level of knowledge in science. For low achieving students this could be problematic in that more weight on effort is allowing them to easily obtain passing grades when in fact they are not "learning" the material.

The factor analysis revealed four types of assessments used by these secondary science teachers: constructed response (projects, essays, presentations, etc), assessments created by the teacher or supplied to the teachers, objective assessments, and the use of major examinations. While objective assessments are used most frequently, there is also a dependence on constructed-response types of assessments. Also, there is a component that appears to separate teachers on whether they design their own assessments or use those provided by publishers or others. Teachers tend to use assessments designed primarily by themselves. Major exams is an independent consideration for teachers. These findings are consistent with a study by Stiggins and Bridgeford (1985) that found science teachers use their own objective tests most often.

This study found that secondary science teachers separate the cognitive level of assessments into two main categories: recall knowledge and higher-order thinking (student reasoning, understanding, and application of material). It appears that for many science teachers there is nearly as much emphasis at the recall level as at understanding. This finding differs slightly from a study by Bol and Strage (1996) that found over half of the teachers' assessment methods required only basic knowledge, while almost none required application. The researchers interviewed these teachers and found that they

actually wanted their students to develop higher order study skills, but did not realize the contradiction between their instructional goals and assessment practices. The findings of the present study differ slightly from an investigation by Doyle (1983) and Gallagher and Tobin (1987) that found little emphasis was placed on applications of scientific knowledge in daily life or on development of higher-order thinking skills. These studies were conducted prior to reform efforts to change the way student learning is assessed in science classes. Therefore, it is not surprising that the present study finds secondary science teachers incorporating more assessment practices that emphasize higher-order thinking skills. Further research is needed to explore the actual tools and methods these science teachers use to assess higher-order thinking skills.

The relationship between grade level and the twelve component scores yielded little differences, with the exception of major examinations. With respect to differences according to ability level of the class, clear patterns emerged. Positive relationships exist between ability level and use of academic achievement, major examinations, and assessment of higher-order thinking, and negative relationships with assessment of recall knowledge and use of graded homework. Higher ability students have an advantage since their teachers tend to use types of assessments, cognitive levels of assessments and grading criteria that mirror that encouraged by recent literature such as use of performance assessments. It appears that low ability students, in contrast, are at a disadvantage since their teachers emphasize recall knowledge and graded homework and focus less on academic achievement and higher order thinking.

While the results of this study are limited by demographics and locations (Virginia is in the midst of a statewide assessment program consisting of all multiple

choice tests, with the exception of writing), comprehensive nature of the sample suggests strong external validity. The responses were based on actual practice, not beliefs, and represented inner city, suburban, and rural schools. Future research on assessment practices may find that the components identified are useful categories for asking questions and relating assessment and grading practices to student motivation and achievement.

References

American Federation of Teachers, National Council on Measurement in Education, and National Education Association (AFT, NCME, NEA). (1990) Standards for teacher competence in educational measurement. Washington, DC:. Author.

Bol, L. & Strage, A. (1996). The contradiction between teachers' instructional goals and their assessment practices in high school biology courses. Science Education, 80(2), 145-163.

Brookhart, S. M. (1994). Teachers' grading: Practice and theory. Applied Measurement in Education, 7, 279-301.

Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. Journal of Educational Measurement, 30, 123-142.

Brookhart, S. M. (1991). Grading practices and validity. Educational Measurement: Issues and Practice, 10, 35-36.

Cizek, G.J., Fitzgerald, Shawn M., & Rachor, R.E. (1996). Teachers' assessment practices: preparation, isolation and the kitchen sink. Educational Assessment, 3(2), 159-179.

Doyle, W. (1983) Academic work. Review of Educational Research, 53(2), 159-199.

Feldman, A. & Alibrandi, M. (1998). Grading with points: The determination of report card grades by high school science teachers School Science & Mathematics, 98(3), 140-147.

Frary, R.B., Cross, L.H. & Weber, L.J. (1993). Testing and grading practices and opinions of secondary teachers of academic subjects: Implications for instruction in measurement. Educational Measurement: Issues and Practice 12(3), 23-30.

Gallagher, J.J. & Tobin, K. (1987). Teacher Management and student engagement in high school science. Science Education, 71(4), 535-555.

Gullickson, A.R. (1985). Student evaluation techniques and their relationship to grade and curriculum. Journal of Educational Research, 79(2), 96-100.

Lawrenz, F. & Orton, R.E. (1989). A comparison of critical thinking related teaching practices of seventh and eighth grade science and mathematics teachers. School Science and Mathematics, 89(5), 361-372.

Marso, R.N. & Pigge, FL (1993). Teachers' testing knowledge, skills, and practices. In S. L. Wise (Ed.) Teacher training in measurement and assessment skills. Lincoln, NE: Burros Institute of Mental Measurements.

Plake, B. S., & Impara, J. C. (1993). Assessment competencies of teachers: A national survey. Educational Measurement: Issues and Practice, 12, 10-25.

Stiggins, R.J. & Bridgeford, N.J. (1985). The Ecology of Classroom Assessment. The Journal of Educational Measurement, 22(4), 271-286.

Stiggins, R. J. & Conklin, N. F. (1992). In teacher's hands: Investigating the practices of classroom assessment, Albany, NY: State University of New York.

Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school: Building a research agenda. Educational Measurement: Issues and Practice, 8, 5-14.

Tittle, C. K. (1994). Toward an educational psychology of assessment for teaching and learning: Theories, contexts, and validation arguments. Educational Psychologist, 29, 149-162.

Table 1

Number and Percent of Teachers by Grade Level and Ability Level of Class

Grade Level	6	7	8	9	10	11	12
	31(12)	43(17)	39(15)	58(23)	51(20)	19(8)	12(5)

Ability Level	AP/Honors	Standard	Basic/Remedial	Mixed
	59(23)	127(49)	23(9)	50(19)

Table 2

Means and Standard Deviations of All Items Measuring Assessment and Grading Practices for Secondary Science Teachers

(n=213)

Factors Used in Determining Grades	Mean	SD
Disruptive student performance	1.63	.96
Improve of performance since the beginning of the year	2.85	1.15
Student effort-how much the student tried to learn	3.25	1.09
Ability levels of the students	3.37	1.32
Work habits and neatness	2.82	1.03
Grade distributions of other teachers	1.27	.69
Completion of homework (not graded)	2.93	1.14
Quality of completed homework (graded)	3.48	1.02
Academic performance as opposed to other factors	4.26	1.08
Performance compared to other students in the class	2.09	1.13
Performance compared to a set scale of percentage correct	4.40	1.30
Performance compared to students from previous years	1.49	.87
Specific learning objectives mastered	4.23	.97
Formal or informal school or district policy of the percentage of students who may obtain As, Bs, Cs, Ds, Fs	1.75	1.23
Degree to which the student pays attention and/or participates in class	3.22	1.09
Inclusion of 0s for incomplete assignments in the determination of final percentage correct	3.82	1.28
Extra credit for nonacademic performance (e.g., bringing in items for food drive)	1.53	.84
Extra credit for academic performance	2.61	1.16
Effort, improvement, behavior and other "nontest' indicators for borderline cases	2.94	1.08

Types of Assessments		
Major exams	2.94	1.11
Oral presentations	2.40	.90
Objective assessments (e.g., multiple choice, matching, short answer)	4.03	.90
Performance assessments (e.g., structured teacher observations or ratings of performance such as a speech or paper)	3.08	.96
Assessments provided by publishers or supplied to the teacher (e.g., in instructional guides or manuals)	2.53	1.11
Assessments designed primarily by yourself	4.36	1.11
Essay-type questions	3.22	.92
Projects completed by teams of students	2.98	1.04
Projects completed by individual students	3.31	.94
Performance quizzes	3.69	.79
Authentic assessments (e.g., "real world" performance tasks)	2.88	1.03
Cognitive Level of Assessments		
Assessments that measure student recall knowledge	3.55	.82
Assessments that measure student understanding	4.21	.77
Assessments that measure student reasoning (higher order thinking)	3.95	.84
Assessments that measure how well students apply what they learn	4.04	.86

Table 3

Percentages of Secondary Science Teachers' Responses for Factors Used in Determining Grades, Types of Assessments Used, and Cognitive Level of Assessments

(n=261)

	Not at All	Very Little	Some	Quite a Bit	Extensively	Completely
Factors Contributing to Grades						
Improvement of performance since the beginning of the year	15	20	39	20	5	2
Student effort – how much the student tried to learn	7	15	36	32	6	3
Ability levels of the students	12	14	25	29	17	4
Assessments that measure student reasoning	0.0	1.7	29.6	41.7	21.7	5.2
Performance compared to other students in the class	38	28	19	11	3	2
Performance compared to a set scale of percentage correct	4	5	15	21	35	21
Types of Assessments Used						
Performance assessments	4	18	45	25	7	1
Authentic assessments	10	21	43	20	4	1
Assessments designed primarily by yourself	1	2	23	27	33	14
Cognitive Level of Assessments						
Assessments that measure student reasoning	0	2	28	43	23	4

Table 4

Component Loadings for Grading Factors, Types of Assessments, and Cognitive Levels of Assessments

	Components											
	Grading 1	Grading 2	Grading 3	Grading 4	Grading 5	Grading 6	Type 1	Type 2	Type 3	Type 4	Level 1	Level 2
Grading Factors												
Student effort – how much the student tried to learn	**.812**	.070	.182	.015	.115	.031						
Improvement of performance since the beginning of the year	**.784**	.145	.033	-.048	-.038	.155						
Ability levels of the students	**.549**	.133	.113	.062	.067	.028						
Work habits and neatness	**.519**	.123	.178	-.096	.318	.409						
Performance compared to students from previous years	.176	**.821**	.007	.027	-.095	.085						
Grade distribution of other teachers	.056	**.581**	-.001	-.246	.061	.291						
Performance compared to a other students in the class	.155	**.570**	.061	-.037	.063	-.030						
School or district policy of percentages of students receiving different grades	.062	**.480**	.116	-.067	.127	-.085						
Extra credit for nonacademic performance	.009	.159	**.615**	.048	.077	-.087						
Extra credit for academic performance	.125	.058	**.615**	-.043	-.302	.382						
Effort, improvement, behavior and other "nontest" indicators for borderline cases	.231	.022	**.533**	.072	.136	.052						
Performance compared to a set scale of percentage correct	-.103	-.070	.068	**.613**	.009	.068						
Academic performance as opposed to other factors	.193	-.038	-.136	**.576**	-.038	-.046						
Specific learning objectives mastered	.272	-.050	-.058	**.462**	.072	.131						

Degree to which the student pays attention and/or participates in class	.344	.237	.259	.004	**.611**	.042						
Quality of completed homework	.178	.014	.086	.273	.020	**.464**						
Completion of homework not graded	.072	.001	.293	.043	.198	.073						
Disruptive student behavior	.015	.302	.197	-.310	.264	-.087						
Inclusion of zeros for incomplete	-.159	-.051	-.019	.348	.005	.079						
Types of Assessments												
Performance assessments (e.g. structured teacher observations or ratings of performance such as a speech or paper)							**.725**	.216	.098	-.217		
Projects completed by teams of students							**.715**	-.085	-.195	-.102		
Oral presentations							**.702**	-.021	-.247	.180		
Projects completed by individual students							**.674**	.183	.127	-.377		
Essay-type questions							**.571**	-.210	.025	.315		
Authentic assessments (e.g. "real world" performance tasks)							**.522**	.048	-.105	.276		
Assessments designed primarily by yourself							.279	**-.749**	.400	-.187		
Assessments provided by publishers or supplied to the teacher							.064	**.855**	-.182	.245		
Objective assessments							-.012	.397	**.760**	-.034		
Performance quizzes							.248	.244	**.678**	.158		
Major exams							.041	-.274	.171	**.772**		
Cognitive Level of Assessments												
Assessments that measure student reasoning (higher order thinking)											**.899**	-.203
Assessments that measure student understanding											**.875**	.128
Assessments that measure how well students apply what they learn											**.852**	-.246
Assessments that measure student recall knowledge											.298	**.940**
% of variance accounted for	20.67	12.32	8.57	7.68	5.98	5.31	24.80	15.64	12.61	10.22	59.70	25.03

Table 5

Statistically Significant Component Score Differences by Grade Level[1]

	Grade Level								
Component	6 n=33	7 n=43	8 n=39	9 n=60	10 n=53	11 n=20	12 n=12	F	p<
Major exams	-.30	-.61	-.40	.27	.48	.24	.42	8.55	.000
Higher order thinking	-.20	-.15	.42	-.09	-.14	.31	.28	2.35	.032
Recall knowledge	-.16	.05	-.13	.21	.17	-.20	-.74	2.20	.043

[1] Component scores are normalized with a mean of 0 and a standard deviation of 1.

Table 6

Statistically Significant Component Score Differences by Ability Level of Class[2]

	Ability Level of Class				
Component	Advanced/AP n=59	Standard n=126	Basic/Remedial n=23	F	p<
Academic achievement	.27	.00	-.22	3.40	.021
Graded homework	.12	-.07	-.38	3.80	.011
Major exams	.24	.05	-.14	3.67	.013
Higher-order thinking	.47	-.08	-.32	6.33	.000
Recall knowledge	-.40	.09	.45	5.10	.002

[2] Component scores are normalized with a mean of 0 and a standard deviation of 1.